

T0B121"0002001

Figure 1-A

*T. reesei bgl4* nucleotide sequence (non-coding is in bold, coding is in regular font)

TTATAGTCGCTTTGTTAAATTGGCCCTCGAGGTCGACCCACGCGTCCGGCTGCTTGTCCCGTTCTGTGCTGATGTCTATC  
TGCCGTTGGCCTCCTCATCTCCTGTTGTCTGTCTCCTCTTAGTGCTTCACTGACGCTAGGTTGGTCACTTTGT  
CCCCCTTCGTTGCTGTGTTCAAGGTCTACCTGCAGTGGTTTGAACCTTGATATCTCTGCTTGAGCATCCGCGT  
CGCCATATAGAGCAGCATATCTTCTATCTCCAAAGATCCCCCTCACCGAGAGTTCTATTCAACCCGACCCCTTGCTGTCAT  
CCAGTCTTCCATCATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCTGGCCGAGAAAGTCTGCTGCTGG  
CTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCCTCTCTCGCTTTACAGATGGCCCCAACGGCG  
TAAGAGGACCAAGTTCTTCAATGGCGTCCCTGCGGCTGCTTCCCTTGGCGCACGTCGCTCGGTTCCACATTCAACCAAA  
CTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAGAGGCCATCGCTAAGATGCGCATGTGATCCTCGGCCGACTATCA  
ACATGCAACGCTCCCTCTCGGTGACGTGGCTTCGAGTCGATTTGGTGGAGTCCGTTCTTGGCGGCTTGGGAGCTGCGG  
CTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTGTGCAATGATCAGGAGGACAGGCGCA  
TGATGGTGCAGAGCATCGTACGGAGCGGCTCTCCGTGAAATCTACGCACTCCGTTCCAGATTGCTGTGCGAGACTCCC  
AGCCGGGTGCGTTTCATGACGGCGTACAATGGCATCAATGGCGTGTGTCAGCGAGAACCTTAAATATCTTGTATGGGATGC  
TTCGAAAGGAATGGGTGGGATGGCTTAATCATGAGGACTGTTACGGCACATACAGTACCACAGAACCGCTTGTGGCAG  
GCCTCGACCTCGAGATGCCCGGACCTCCACGCTTCGAGGAGAAACACTCAAAGTTCAACGTCTCCAACGGAAAGCCCTTTA  
TCCACGTCAATTGACAGAGGCTAGGAAAGTTCTTCAGTTCGTCAAGAAAGTGTGCTGCTCCGAGTGAACGAGAACGGCC  
CCGAGACGACTGTCAACAAACACCCCCGAAACGGCAGCTCTCCCTCCGGAAGTTGGCAACGAGGGCATCGTGTGCTGAAGA  
ACGAGAACAAACGTTCTGCCCTTGAGCAAGAAAGAACGCTGATTGTCCGCCCCAACGCCAAGCAGGCCACATACCCAG  
GCGGAGGCTCTGCCGCACTCAGGGCTACTACGAGTCACTCCCTTTGACGGCTCAGCAAGCAGCTCGAGACGCCGCCAT  
CGTACACCGTCGCGCCTACACCAACCGTTCTCTCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGCGCTCCGGGCATGC  
GCTGGAGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCCGACACATTGACGAGCTCTTCTTCAACCAAGACGACATGC  
ACCTGGTGACTACTACACCCCAAGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACT  
GCACCTACGAGCTCGGCTCGTGTGCGGCACGGCAAGCGGTACGTAGACGACCAAGCTCGTCTCGACCAACGCCACCA  
AGCAGGTCCCCGGCGATGCTTCTTCGGCTCCGCCACCCCGGAGGAGACGGGCGCATCAATCTCGTCAAGGGCAACACGT  
ACAAAGTTCAAGATCGAGTTCGGTCCGCAACCCACCTACACCTCAAGGCGACACCATCGTCCCCGGCCACGGCTCCCTCC  
GCGTCGGCGGCTGCAAGGTTCATTGACGACCCAGGCCGAAATCGAAAGTCCGTGCCCCCTCGCCAAAGGACGACCAAGGTCA  
TCATCTGCGGGGCTTAACGCCGACTGGGAGACCGAGGGCCGACCCGCGGAGCATGAAGTCCCCCGGCGTGTGGACC

TTTTTTT"0004200T

Figure 1-B

*T. reesei* *bgl4* nucleotide sequence (non-coding is in bold, coding is in regular font)

AGCTCATTGCCGACGTGGCCGCGGAAACCCAAACACCGTCGTCTGTCATGCAGACGGGCACCCCGAGGAGATGCCCTGGC  
TCGACGCCACGCCCGCGTCATCCAGGCCCTGGTACGGGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCCTTTGGCG  
ACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCTGCAGGACAAACCCCGGTTTCTCAACTTCCGCACCG  
AGCCGGGCGCACGCTGTACGGCGGAGGACGTCACGTCGGGTACAGGTACTACGAGTTTGGCGACAAGGACGTCGAATTTCC  
CCTTTGGCCACGGCCTGTCTACACCACTTTTGGCCTTTTCCAATCTCTCCGTGTCTACAAGGACGGCAAGCTGAGCGTGT  
CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGCGCACAGGTGGCCAGCTCTACGTCAAGCCCTCCAAGCGGCCAAGA  
TTAACCGCCCGTCAAGGAGCTCAAGGGCTTCGCAAGGTCGAACTGCAGCCCGCGAGACGAAAGCGGTGACAATCGAGG  
AGCAGGAGAAAGTACGTCGCTGCGTATTTTGTATGAGGAGCGGGATCAGTGGTGTGTCGAAAAGGTGACTATGAGGTTATCG  
TGAGCGACAGCAGCGCAGGATGGGTTGCGCTCAGGGTAAGTTTACGGTGGGAGAGACGTATTGGTGGTCTGGCG  
TGTAAAGTCTGTCATCTTTGGCAGATTGAATCCAGTCACTTTGCATATAGCGCCGATGAAATGAGAAACCAACGATCAT  
TGTGTAATAA

60

75  
136

61  
76

TCSTT"0004200T

Figure 2-A

Predicted *T. reesei* BGL4 amino acid sequence

MADIDVEAIL	10	20	30	40	50	60	70
KKLTIAEKVD							
LLAGIDFWHT							
KALPKHGVPS							
LRFTDGPNGV							
RGTKFFNGVP							
AACFP CGTSL							
GSTFNQTLLE	80	90	100	110	120	130	140
EAGKMMGKEA							
IAKSAHVILG							
PTINMQRSPL							
GGRGFESIGE							
DPFLAGLGAA							
ALIRGIQSTG							
VQATIKHFLC	150	160	170	180	190	200	210
NDQEDRRMMV							
QSIIVTERALR							
EIYALPFQIA							
VRDSQPGAFM							
TAYNGINGVS							
CSENP KYLDG							
MLRKEGWGDG	220	230	240	250	260	270	280
LIMSDWYGTY							
STTEAVVAGL							
DLEMPGPPRF							
RGETLKFNV							
NGKPFHVID							
QRAREVLQFV							
KKCAASGVTE	290	300	310	320	330	340	350
NGPETTVNNT							
PETAALLRKV							
GNEGIVLLKN							
ENNVLP LSKK							
KKT LIVGPNA							
KQATYHGGGS							
AALRAYYAVT	360	370	380	390	400	410	420
PFDGLSKQLE							
TPPSYTVGAY							
TTVPPILGEQ							
CLTPDGAPGM							
RWRVFNEPPG							
TPNRQHIDEL							
FFT KTDMHLV	430	440	450	460	470	480	490
DYYHPKAADT							
WYADMEGTYT							
ADEDCTYELG							
LVVC GTAKAY							
VDDQLVVDNA							
TKQVP GD AFF							
GSATREETGR	500	510	520	530	540	550	560
INLVKGNTYK							
FKIEFGSAPT							
YTLKGDTIVP							
GHGSLRVGCG							
KVIDDQAEIE							
KSVALAKEHD							
QVIICAGLNA	570	580	590	600	610	620	630
DWETEGADRA							
SMKLP GVLDQ							
LIADVAAANP							
NTVVVMQTGT							
PEEMPWLDAT							
PAVIQAWYGG							

FOBT2T" 0002200T

Figure 2-B

Predicted T. reesei BGL4 amino acid sequence

NETGNSIADV	VFGDYNPSGK	LSLSFPKRLQ	DNPAFLNFR	T EAGRTL	YGED	VVGYRY	YEF	ADKDVNFPFG	
640	650	660	670	680	690	700			
HGLSYTTFAF	SNLSVSHKDG	KLVSLSVKN	TGSVPGAQVA	QLYVKPLQAA	KINRPVKELK	GFAKVELQPG			
710	720	730	740	750	760	770			
ETKAVTIEEQ	EKYVAAAYFDE	ERDQWCVEKG	DYEVIVSDSS	AAKDGVALRG	KFTVGETYWW	SGV			
780	790	800	810	820	830	833			